R E P O R T R E S U M E S

ED 017 972 24 EA 001 074

PROBLEMS IN THE MEASUREMENT AND EVALUATION OF STUDENT DISCUSSION GROUPS.

BY- GRIB, THOMAS F.
SAINT NORBERT COLL., WEST DE PERE, WIS.

REPORT NUMBER BR-5-0923 PUB DATE 2 SEP 66
CONTRACT OEC-6-10-244
EDRS PRICE MF-$0.25 HC-$0.56 12P.

DESCRIPTORS- #COLLEGE STUDENTS, #DISCUSSION GROUPS, #MEASUREMENT TECHNIQUES, EVALUATION CRITERIA, #TEACHING METHODS, TAXONOMY, TEACHER RATING, #RESEARCH PROBLEMS, ACADEMIC ACHIEVEMENT, INSTRUCTIONAL IMPROVEMENT, COGNITIVE PROCESSES, WEST DE PERE,

[redacted] PAPER EXAMINES SOME OF THE MEASUREMENT AND EVALUATION PROBLEMS ENCOUNTERED IN RESEARCH ON STUDENT-LED SMALL GROUP DISCUSSION TEACHING PROCEDURES. IT FOCUSES UPON PROBLEMS OF CRITERION EXAMINATION, USE OF RATING FORMS TO ASSESS THE EFFECTIVENESS OF VARIOUS METHODS OF INSTRUCTION, AND METHODS FOR CHARACTERIZING CLASSROOM PROCEDURE. THIS PAPER WAS PRESENTED AT THE AMERICAN PSYCHOLOGICAL ASSOCIATION CONVENTION (NEW YORK CITY, SEPTEMBER 2, 1966). (AUTHOR/DG)

DOCUMENT FILMED FROM BEST AVAILABLE COPY

ED017972

Paper read at American Psychological Association Convention
September 2, 1966. New York

BR-5-0723

Problems in the measurement and evaluation of student discussion groups[1]

PA 24

Thomas F. Grib
St. Norbert College

This paper will examine some of the problems of measurement and evaluation we encountered in our research on student-led discussion procedures. In a few cases it offers some suggestions for solutions to these problems, but its primary purpose is to point out what we have found to be the main difficulties so that others beginning research in this area may approach the problems with more awareness of them than we had.

Criterion examinations

A general problem in the evaluation of teaching methods is the choice of a criterion or measure of success. In educational research this usually includes some form of comprehensive final examination.

Two questions arise here: (1) Does the criterion examination faithfully measure the objectives of the new teaching method?

(2) If the criterion measures do have a well-founded relationship to the objectives of the new procedure, are they then biased in favor of the new method?

Let me take up the second question first.

It is rare that two different teaching procedures have exactly the same objectives. The problem is then how to develop an examination that is equally fair to the two (or more) different teaching procedures. My opinion is that this is an extremely difficult, if not impossible, task to accomplish.

---

1. This work was supported in part by the Cooperative Research Branch, United States Office of Education, Contract No. O.E.-6-10-244.

EA 001 074

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Good research procedure requires that we employ a control group, differing from the experimental group in no way except that it does not receive the experimental treatment. But are the experimental and control groups taught by methods that have precisely the same objectives? If not, why compare one with the other? Perhaps we should concentrate on parametric rather than comparative studies. In a comparative study we evaluate one method in comparison to another alternative method. Thus our measure of effectiveness is relative and the outcome is entirely dependent upon the choice of the alternative method. Furthermore, in a new area of research we may be jeopardizing our venture by jumping into comparative studies prematurely. When a new method of teaching is being compared to the conventional method the new procedure may be at a disadvantage in two ways. First, the new method may not be developed to its full potential and, second, the teacher using the new procedure is likely to be inexperienced in its use.

On the other hand, in a parametric study we would explore the functional relationships and interactions of various aspects of the experimental variable. By doing this we could discover the most favorable combination of these variables that lead to the best results. To date we have made no systematic inquiry into the relative gains from variations in: the role of the teacher, student leadership, length and frequency of discussions, the sequencing of discussions, variations in feedback meetings, type and author of guide questions, etc.

Now let us turn to the first question: Does the criterion examination faithfully reflect the objectives of the new teaching method?

In our case the "new" method was the substitution of small discussion groups comprised of 4-6 students for the usual lecture or professor-led discussion. We felt the smaller student-led discussions would increase the motivation and responsibility of the student for his own learning, force active rather than passive participation, and require the organization and verbalization of learned

material. This, in turn, would lead to increased comprehension of the course materials, a shift of emphasis from memory and recall to understanding, the development of critical and analytical thinking, and an increased ability to apply learned methods and principles to problem solving situations. These, then, were the general objectives of the method.

These objectives are related to certain processes we intended to develop in the student by the discussion method which we hoped would be somewhat independent of course content.

Our problem was to construct examinations that would measure these objectives. We decided to use Bloom's Taxonomy of Educational Objectives, (Bloom, 1956) as the basis for classifying test items on the criterion examinations.

This was a big problem, mainly because we did not know what prior experiences the student brought with him to the exam.

The developers of the Taxonomy decided that the basis for their classification of educational objectives would be the student behavior which a test item is intended to elicit. In reaching this decision they chose to disregard the student behavior which an item actually evokes. In doing this, they acknowledged that the behavior which an item actually evokes and that which it is intended to evoke may be different due to prior experiences of the examinees.

In our own research we attempted to classify the test items on the final examinations according to the Taxonomy. We hoped to show that the discussion procedure would result in better performance at the higher levels of the Taxonomy, i.e., from comprehension through evaluation. Conversely we felt that the lecture method would be at least as good as, if not better than, the discussion method for relaying information.

Unfortunately, there was not enough time to construct new examinations following the Taxonomy and so, in most cases, the examinations were revised forms of final examinations given in prior years. After these examinations were

administered, the project staff attempted to classify the test items according to the Taxonomy. It became immediately apparent that this was not an easy job, mainly because we were unable to ascertain what prior experiences the students brought with them to the examinations.

Rather than give up the attempt to classify items we decided to rate the items at the lowest level at which they could be answered by a student. If, for example, an item was written with intent to elicit comprehension by the student but could be answered by merely repeating what he had heard in class (i.e., mere recall), then the item would be classified as knowledge, the lowest level in the Taxonomy.

By following this procedure, we found that most examination items fell into the knowledge category, with a few items classified as comprehension and application. It was not surprising, then, that we found few significant differences between groups using the discussion technique and those taught by conventional methods.

The significant differences we did find were in the course in psychological statistics, where it was easier to write items demanding comprehension and application on the part of the student.

This leads us to another, and perhaps more significant problem. The student discussion techniques were designed to develop critical, analytic thinking, but we still test the student mainly on his knowledge of the content of the course. In other words, we are attempting to develop certain <u>processes</u> in the student such as comprehension, application, analysis and synthesis, but our tests were ineffectual in measuring these behaviors, and concentrated primarily on the course <u>content</u>.

Is it possible, or even desirable, however, to measure a cognitive process independent of content? Kropp and Stoker (1966) conducted a three year research project on the construction and validation of Taxonomy type tests, and one of

their hypotheses dealt with the transcendance of cognitive processes over content. They investigated this hypothesis by factor analysis and found the majority of factors extracted to be mixtures of both process and content.

Their conclusion was that they hypothesis was neither proved nor disproved by their data.

I would like to make some other observations before we leave the problem of the criterion examination. These are not directly related to problems of measurement but are relevant to student behavior.

The first observation, which I'm sure we're not the first to make but we think bears repeating, is that students study for what they're tested on. We can list many objectives in our course syllabus but the students soon learn what they are tested for and study for the tests, not for the stated course objectives.

In our research we have learned that to the extent the student discussions become an exercise (i.e., not related to course grades) the students treat them as an exercise. In our experience, it is not unusual to ask students to do one thing (e.g., comprehension, understanding, application, evaluation, etc.) and then test them on another (recall). Thus, we frequently have a situation where the course objectives and course examinations are working at cross purposes.

Rating Forms

From past experience with student-led discussions we felt that there were changes taking place in the students that we were unable to measure by examinations. We attempted to check on these findings by developing various rating forms and questionnaires to assess the students' opinions regarding various aspects of the discussion method and their own use of the method. Neil Webb (Webb, 1966) has already summarized the results of the End of Course Questionnaire.

In figure 1 we have the results of one administration of the Instructional Method Rating Form, a form which was used to compare student response to different

methods of instruction. This example compares the same students' ratings of a lecture and student discussion on the same material, in this case the binomial theorem. You will note that for each item the discussion method had a higher median rating than did the lecture. (Of course, these results may be interpreted to mean that it was a poor lecture rather than a good discussion.)

Characterizing classroom procedure

Before I finish I would like to take up one more topic and that is the problem of characterizing classroom procedure. Our research contrasted the discussion method with the instructor's "usual approach" or "conventional method." Obviously the terms "usual approach" and "conventional method" are so vague as to be almost meaningless. We wanted to classify each teacher's usual approach in an operational and, if possible, quantitative way.

We began by tape recording the instructor's classes and listening to samples of these. We then set up a classificatory scheme modified somewhat from Flanders and Amidon (Flanders, 1965; Amidon, 1966). The resulting categories are shown in Table 3. The two main categories are teacher talk and student talk; these are then further subdivided according to the nature of the communication.

We listened to random samples of two entire classes for each instructor and recorded the frequency and amount of time for each category. The amount of time spent in each category was then converted into percentages and put in tabular form. An example of the analysis of one class is shown in Table 1. Here the comparison is between a lecture and an instructor-led discussion in psychological statistics. As might be expected there are shifts in the amount of time spent in each category. In the lecture, teacher talk accounted for 95% of the time, but dropped to 65% during the instructor-led discussion. Student talk increased from 5% during the lecture to 35% during the discussion.

You will note that this is not a process type of recording, but merely a summary breakdown of the amount of time spent in various activities. We felt

this was sufficient for our purposes as a means for characterizing classroom procedures.

We have made this type of analysis for professor lectures, professor-led discussion and student-led discussions and are now in the process of comparing them. Obviously this is a rather crude method of analysis, but we feel it is still better than using terms such as the instructor's "usual approach" or "conventional procedure."

These, then, were some of the problems of measurement and evaluation we encountered in our research. I am sure there are some obvious solutions to some of these problems that we have overlooked and I hope that you will suggest them to us during the discussion period.

References

Amidon, Edmund, Interaction Analysis: Recent Developments. (Paper read at the annual meeting of the American Educational Research Association, February, 1966, Chicago, Illinois.)

Bloom, B. S. (ed) Taxonomy of Educational Objectives: Handbook I: Cognitive Domain. New York: David McKay, 1956

Flanders, N. A. Teacher Influence, Pupil Attitudes and Achievement. Washington: U. S. Office of Education, 1965. (Cooperative Research Monograph No. 12)

Kropp, R. W. and Stoker, The Construction and Validation of Tests of the Cognitive Processes as Described in the Taxonomy of Educational Objectives. Tallahassee: Florida State University, 1966. (Cooperative Research Project No. 2117)

Webb, N. J. Innovation in Student Discussion and Related Procedures. (Paper read at annual convention of The American Psychological Association, September, 1966, New York.)

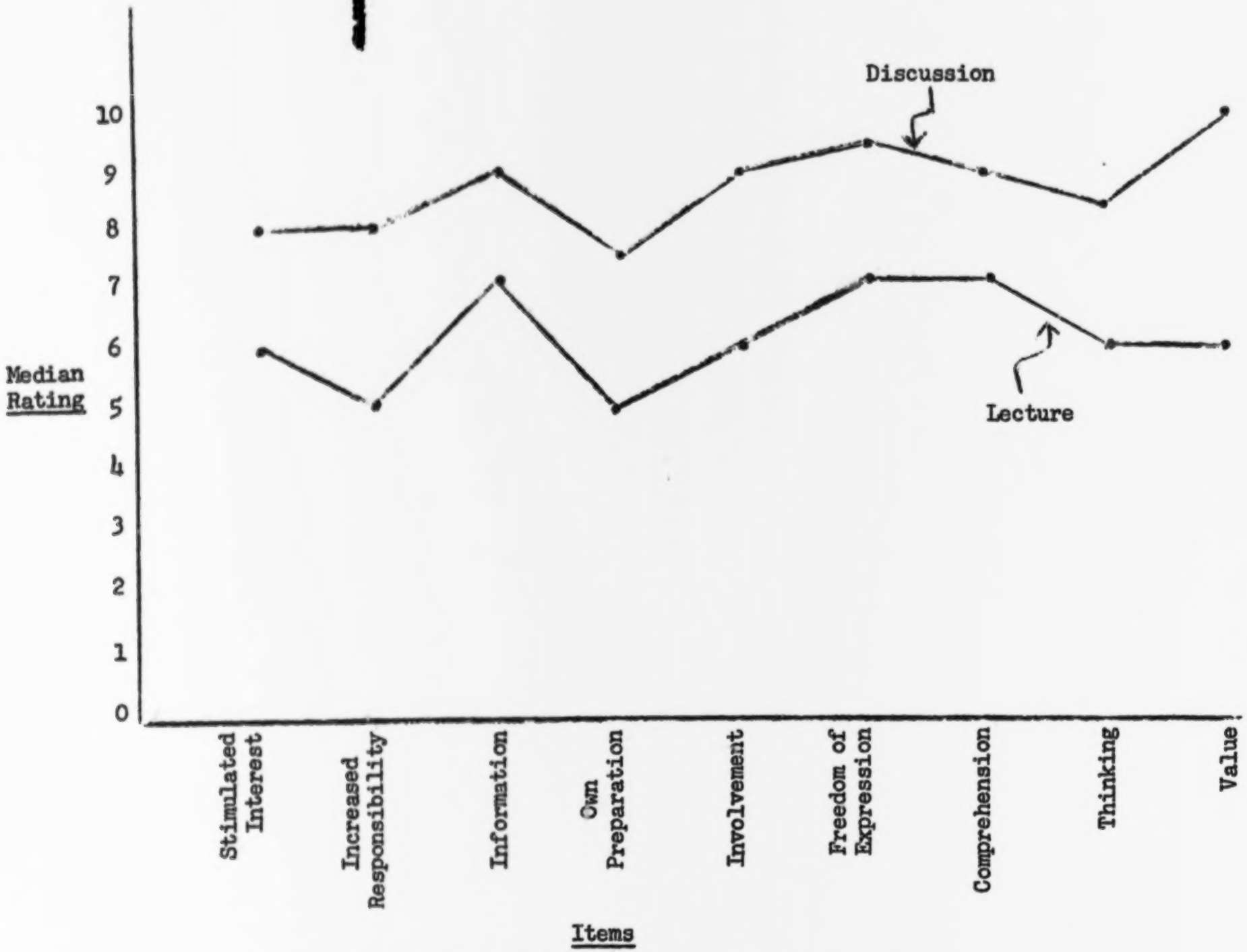


FIG. 1 - Comparison of ratings of lecture vs. discussion on Instructional Method Rating Form.

TABLE 1

Characterization of classroom procedure: lecture vs. instructor-led discussion.

Category	Lecture			Instructor-led Discussion		
	f	time	% of time	f	time	% of time
A. TEACHER TALK	61	46'51"	95.8	103	29'18"	64.8
1. Giving Directions	3	3'00"	6.1	5	52"	1.9
2. Lecturing	29	37' 3"	75.8	36	15'40"	34.6
a. factual	9	16'12"		5	34"	
b. integrative	20	20'51"		30	14'46"	
c. evaluative	0			1	20"	
3. Asking Questions	23	4' 9"	8.4	51	7'47"	17.2
a. factual	19	2' 2"		27	3' 9"	
b. integrative	4	2' 7"		21	4'23"	
c. evaluative	0			3	15"	
4. Answering Questions	6	2'37"	5.3	11	4'59"	11
a. factual	1	7"		3	59"	
b. integrative	4	2' 2"		8	4'00"	
c. evaluative	1	28"		0		
B. STUDENT TALK	29	2' 3"	4.1	71	15"54"	35.1
1. Answering Questions	20	1'25"	2.9	54	13'53"	30.7
a. factual	17	1'16"		25	3'59"	
b. integrative	3	9"		29	9'54"	
c. evaluative	0			0		
2. Asking questions	9	38"	1.2	17	2' 1"	4.4
a. factual	3	13"		6	37"	
b. integrative	5	23"		11	1'24"	
c. evaluative	1	2"		0		

Instructional Method Rating Form

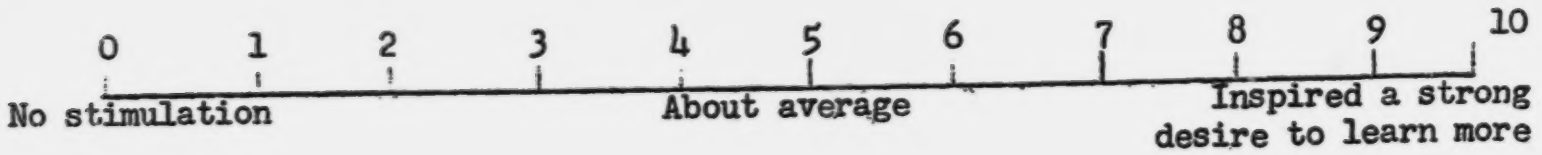
Course Title_____ Instructor_____

Group Letter_____ Date_____

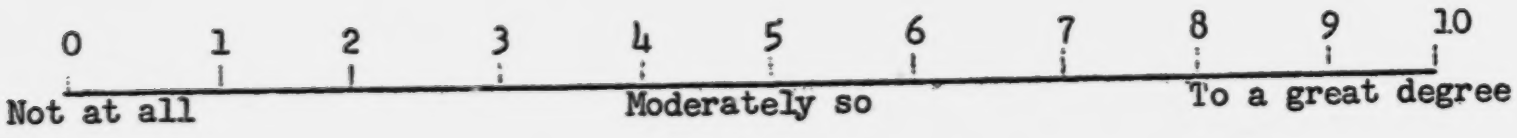
To the student: In this rating form you are asked to make evaluative judgments about today's class. The primary interest in asking you to do this is in order to compare different methods of instruction, rather than characteristics of the teacher.

Each item is rated on a scale which extends from the worst possible condition on the left to the best possible condition on the right. In answer to each item proposed about today's meeting, PUT A CIRCLE around the number that corresponds with your estimate.

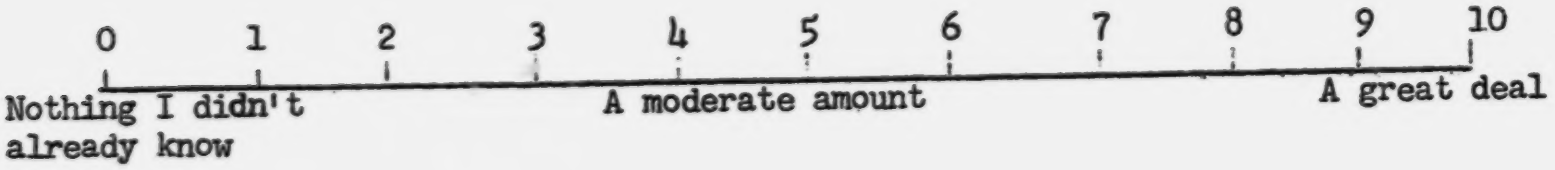
1. How much has today's class stimulated your interest in the course?



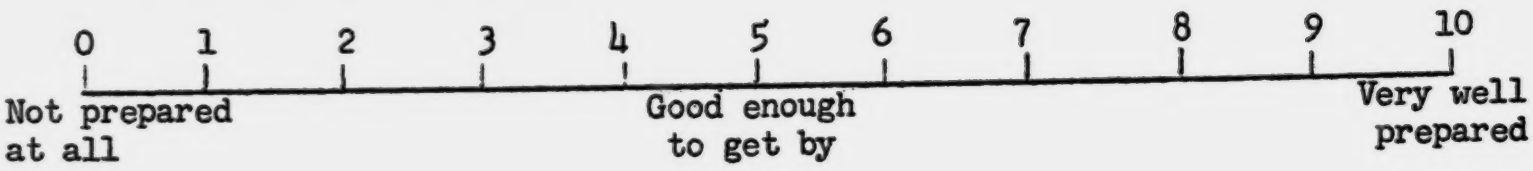
2. How much did today's class stimulate in you a sense of independence and responsibility in your own growth and learning?



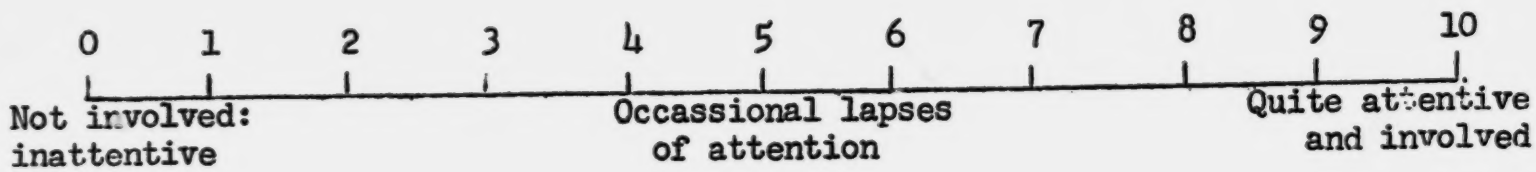
3. How much knowledge or information did you gain in today's class?



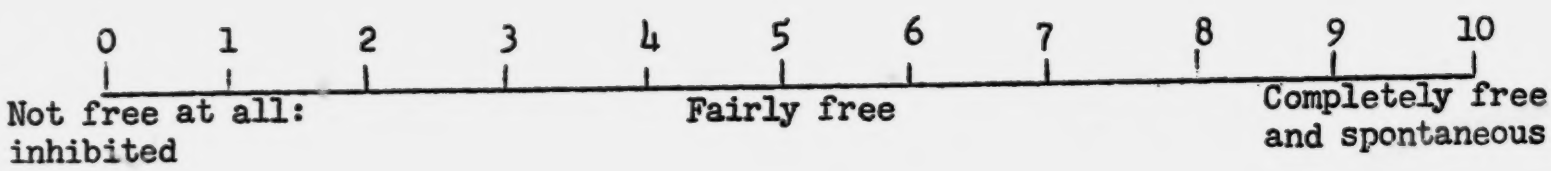
4. My own preparation for today's class was



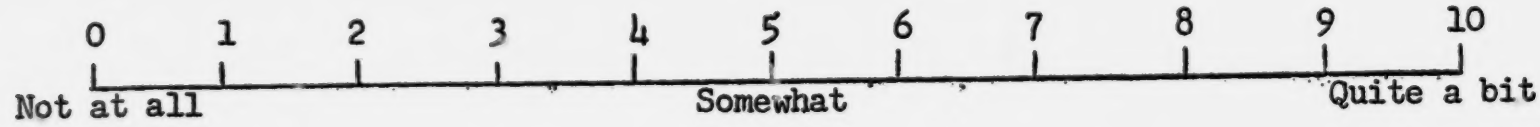
5. How would you rate your own active attention and involvement during today's class?



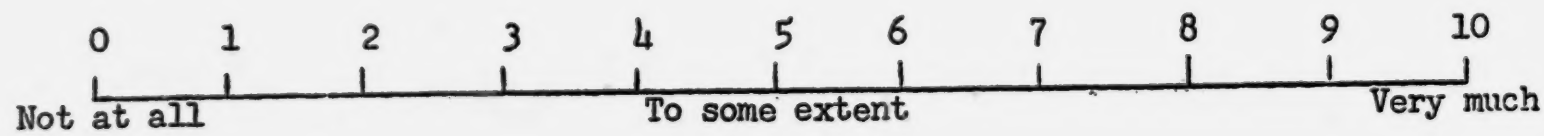
6. How free did you feel in today's class to ask questions, disagree or express your own ideas?



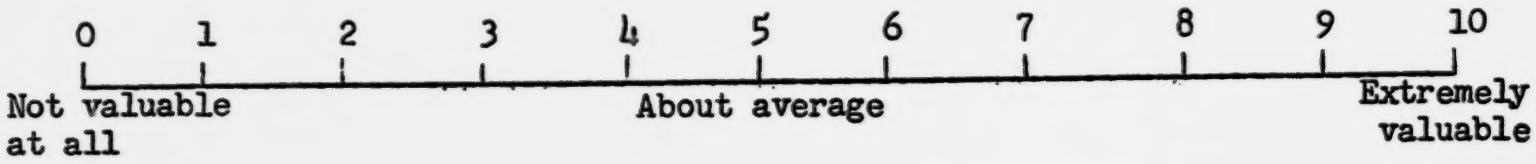
7. How much has today's class pointed out gaps and inadequacies in your comprehension of material?



8. To what extent did today's class encourage critical thinking in the solution of problems?



9. The overall value of today's class for me as a learning experience was



Please use this space to comment on any aspect of today's class or to make suggestions for improvement.